Final Project

Yiru Fei

4/19/2020

Introduction

In this project, I estimated the Rt, the effective reproduction number. This number means the number of people who become infected per infectious person at time t. R0 is the basic reproduction number of an epidemic. If R0 is greater than 1, the epidemic spreads quickly. If R0 is less than 1, the epidemic disappears before everyone becomes infected. And the flu has an R0 between 1 and 2. Measuring Rt can let us know when we might loosen restrictions. If we are able to reduce Rt to below 1, we can reduce the number of new cases and virus becomes manageable.

Poisson Distribution

The first step is that I assume the infection model is poisson distribution, and λ represents the average rate of infections per day, then k represents the new cases on a day. So the function is given by

$$P(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

The distribution of λ over k, the likelihood function, is given by

$$L(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

According to the paper, Real Time Baysian Estimation of the Epidemic Potential of Emerging Infectious Diseases. Based on the standard epidemic susceptible-infected (SIR) model, γ is the infectious period, the relationship between R_t and λ is given by

$$b(R_t) = e^{\tau\gamma(R_t-1)}$$

$$\Delta\frac{T(t) - T(t-\tau)}{\tau} = b(R_t)\Delta T(t)$$

$$\lambda = k_{t-1}e^{\gamma(R_t-1)}$$

According to Bayes' rule, the probability distribution of R_t is

$$P(R_t|k_t) = \frac{P(R_t)L(k_t|R_t)}{P(k_t)}$$

where $L(k_t|R_t)$ is likelihood function in terms of R_t.

If we use $P(R_{t-1}|k_{t-1})$ as prior, and $P(R_t|k_t)$ as posterior, then we can get the equation as

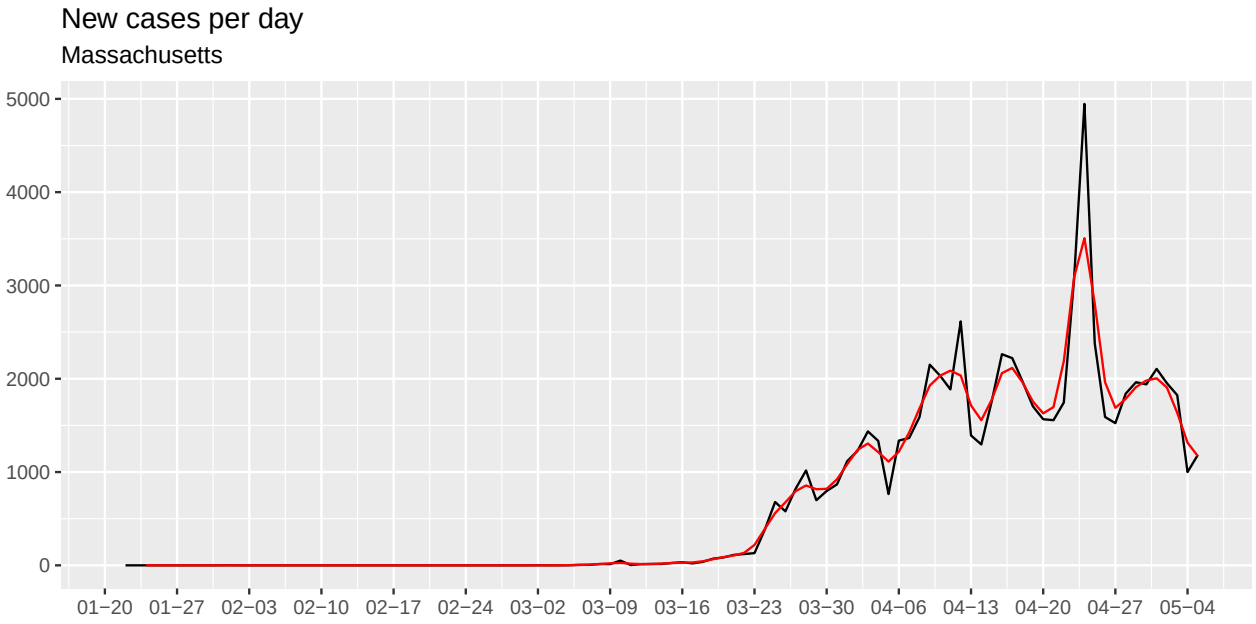
$$P(R_t|k_t) \propto L(k_t|R_t) * P(R_{t-1}|k_{t-1})$$

If we assume t = 0, then $P(R_t|k_t) \propto \prod_{t=0}^{T} L(k_t|R_t) * P(R_0) = \prod_{t=0}^{T} L(k_t|R_t)$

Data Description

I used the US COVID-19 Daily Cases with Basemap for the US from Harvard Dataverser. It contains state and county-level data, but in this project I only used the data for New York and Massachusetts. The first step is to compute the number of new cases every day, and smooth it over a rolling window. The smoothing is essential to account for lags pronounced over weekends. I used the gaussian window smoothing function with 5 days of smoothing windows.

	date	new_cases	new_cases_smooth
32	2020-02-21	0	0
12	2020-02-01	1	0
52	2020-03-12	13	13
14	2020-02-03	0	0
53	2020-03-13	15	14
38	2020-02-27	0	0



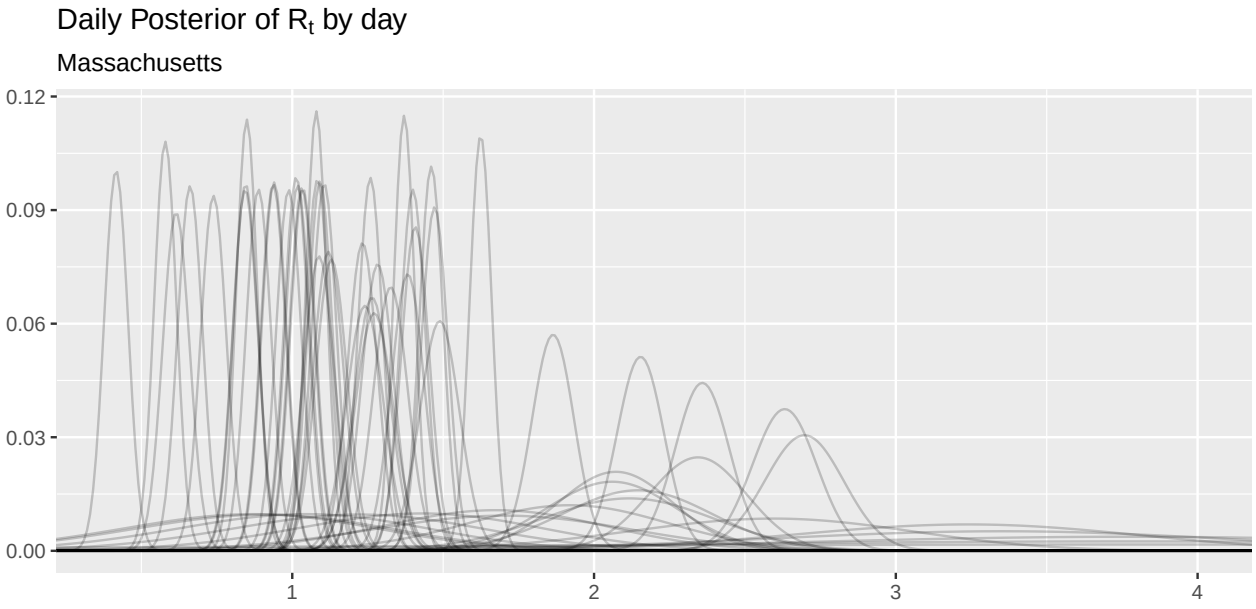
The black line represents daily new cases in Massachusetts, and the red line is the new cases after smoothing. This figure suggests that the number of new cases is increasing every day in Masschusetts, and reached daily new cases maxmimum on April 25th, then the growth rate is gradually slowing.

Calculate Rt

The second step is to compute the likelihoods. I computed log-likelihoods instead of the likelihoods, since it is easier to smooth data over a rolling window. Therefore, when calculating the posterior probabilities, the formula is given by

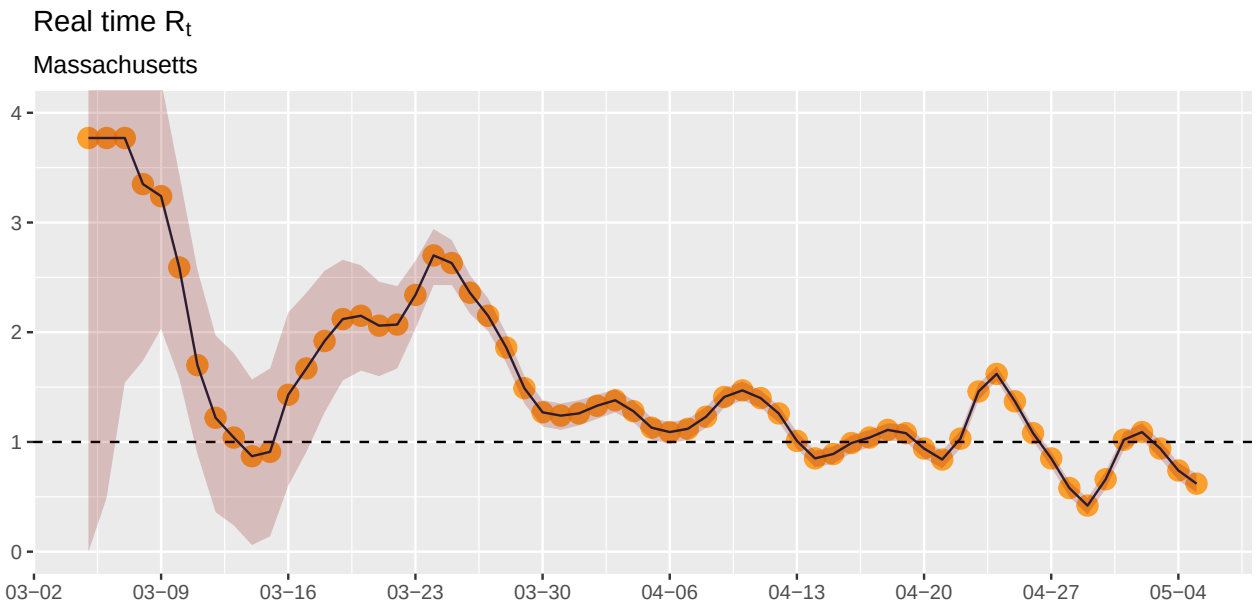
$$P(R_t|k_t) \propto exp[\prod_{t=0}^{T} log(L(k_t|R_t))]$$

date	new_cases	new_cases_smooth	r_t	posterior
2020-04-11	1886	2087	8.52	0
2020-03-18	38	44	6.13	0
2020-05-05	1184	1170	7.06	0
2020-03-16	33	28	5.62	0
2020-04-08	1588	1682	2.17	0
2020-03-29	698	816	8.27	0

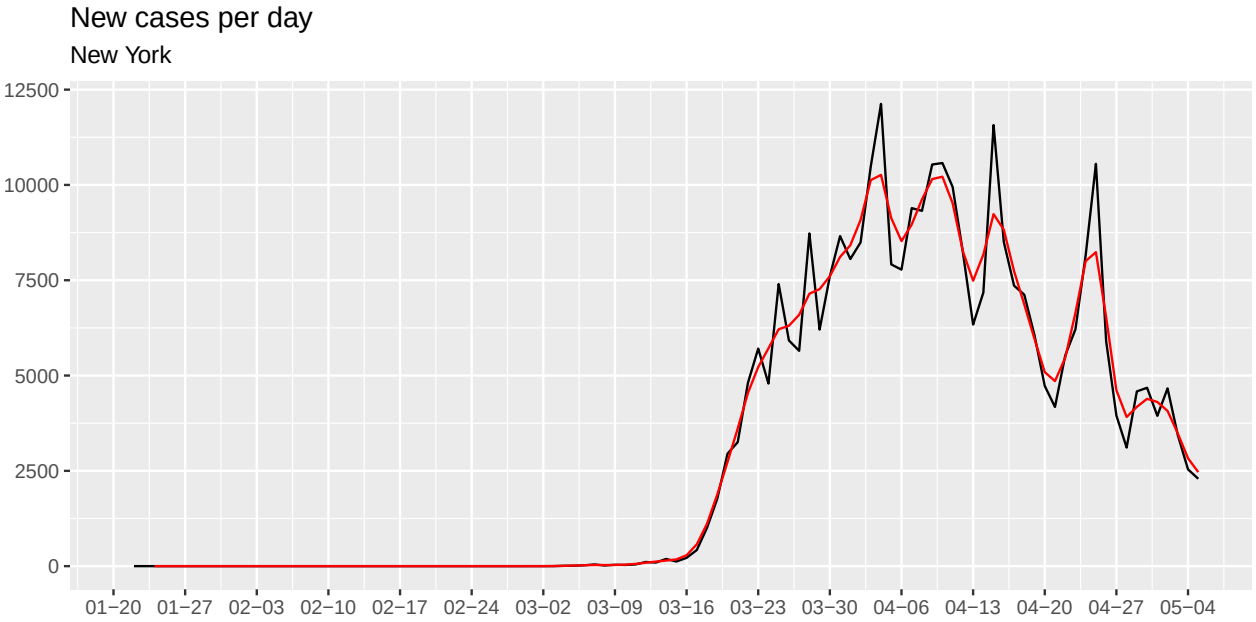


Result

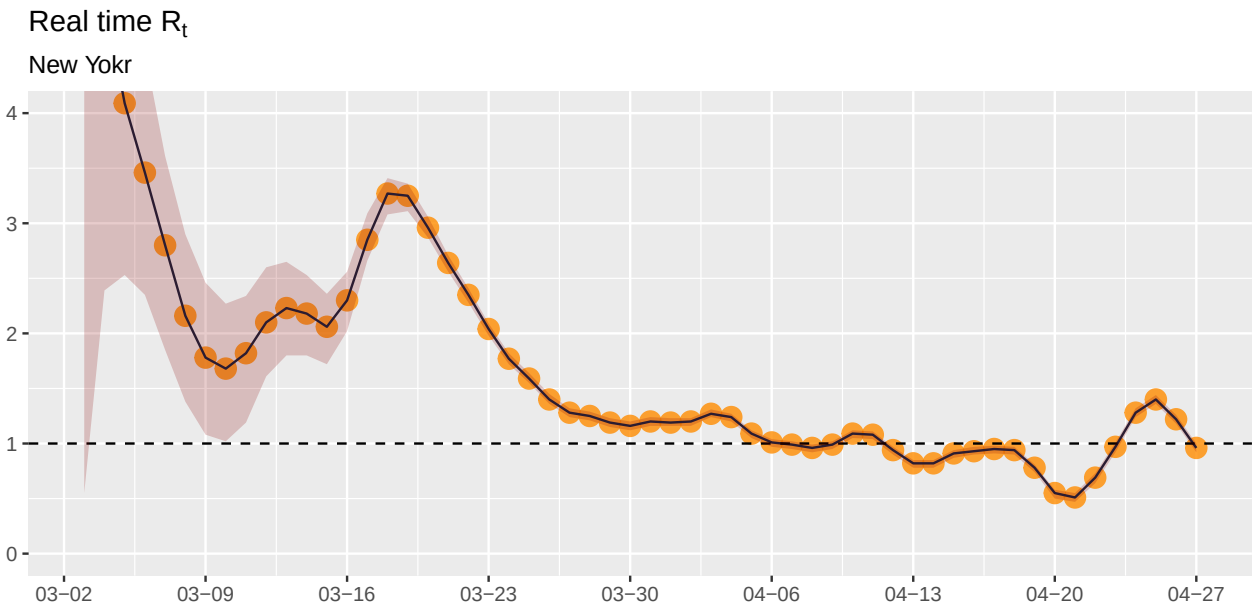
The final step is to estimate the values of R_t and the 95% density intervals surrounding them. Because using Bayesian methods, the model produce highest density interval (HDI). The HDI is wide has to be with the lack of information. Massachusetts seems have reduced Rt, and the gray band close to the orange point suggests that we may conclude that we are below the safety shreshold. After May Rt has dropped below 1, so the virus will slowly disappear in this case.



States Compare



I then explored New York cases, the above plot indicates that the daily new cases in New York show a normal ditribution, and the peak is probably in early April. After comparing with Massachusetts data, I found new cases in New York have increased dramatically a week earlier than Massachusetts, and reached its maximum also a week earlier than Massachusetts. So we may conclued that the infectivity of this virus is basically the same in the eastern United States, and it will take about two months from the outbreak to the end of the infection, which is very similar to the infection period in China. After checking the Rt, I realized the Rt has droped below 1 first time on April 6th, which is still a week earlier than Massachusetts.



Reference

Systrom, Kevin. "The Metric We Need to Manage COVID-19." Systrom, 15 Apr. 2020, systrom.com/blog/the-metric-we-need-to-manage-covid-19/.

"(Tutorial) Estimating COVID-19's in Real-Time (Replicating in R)." DataCamp Community, www.datacamp.com/community/tutorials/replicating-in-r-covid19.

Bettencourt, Luís M. A., and Ruy M. Ribeiro. "Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Diseases." PLoS ONE, vol. 3, no. 5, 2008, doi:10.1371/journal.pone.0002185.